Investigating the reliability and factor structure of Kalichman's "Survey 2: Research Misconduct" questionnaire

A post hoc analysis among biomedical doctoral students in Scandinavia

Holm, Søren; Hofmann, Bjørn

Published in:
Journal of Empirical Research on Human Research Ethics

DOI (link to publication from Publisher):
10.1177/1556264617714658

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Investigating the reliability and factor structure of Kalichman's "Survey 2: research misconduct" questionnaire – a post hoc analysis among biomedical doctoral students in Scandinavia

Søren Holm, Bjørn Hofmann

Abstract

A precondition for reducing scientific misconduct is evidence about scientists' attitudes. We need reliable survey instruments and this study investigates the reliability of Kalichman's "Survey 2: research misconduct" questionnaire . The study is a post-hoc analysis of data from three surveys among biomedical doctoral students in Scandinavia (2010-2015). We perform reliability analysis, and exploratory and confirmatory factor analysis using a split-sample design as a partial validation. The results indicate that a reliable 13 item scale can be formed (Cronbach's Alpha .705), and factor analysis indicates that there are 4 reliable sub-scales each tapping a different construct: 1) general attitude to misconduct (Alpha .768) , 2) attitude to personal misconduct (Alpha = .784), 3) attitude to whistleblowing (Alpha .841), and 4) attitude to blameworthiness/punishment (Alpha .877). A full validation of the questionnaire requires further research. We, never the less hope that the results will facilitate the increased use of the questionnaire in research.

Introduction

Scientific misconduct is increasingly being recognised as a problem across a range of sciences including biomedical science (1-6). The 2010 Singapore Statement on Research Integrity situates practices that constitute research misconduct (fabrication, falsification, plagiarism) within a broader range of irresponsible research practices (7). In survey research on scientific misconduct it is *inter alia* important to be able to quantify the attitudes of respondents towards specific types of misconduct, as well as their general attitude towards misconduct and those that engage in misconduct. If there was a validated scale for this purpose in general use it would greatly improve the comparability of studies and thereby the development of knowledge in the field. There are, however to our knowledge no validated general scales for this purpose. There are a few more specific questionnaires that have been validated. The Scientific Misconduct Questionnaire—Revised (SMQ-R) has been validated, but it is narrow in scope measuring clinical trial coordinators views about, and experiences of misconduct (8). Croatian researchers have validated anattitudes toward plagiarism questionnaire (9). A validation of a 42 item Responsible Conduct of Research Measure is also described in the literature (10), but the questionnaire does not seem to be publicly available. Recently US researchers have published a validation of the How I Think about Research (HIT-Res) measure which measures compliance disengagement in researchers, a construct that to some extent overlaps with the respondents' attitude towards misconduct (11).

We have, in the absence of validated instruments to measure general attitudes towards scientific misconduct used the 2005 version of Kalichman's "Survey 2: research misconduct" questionnaire as a tool to gauge attitudes towards misconduct in a number of studies. The questionnaire is short

which is an advantage in relation to combining it with other measures in a larger questionnaire, but it never the less covers issues of data related misconduct, plagiarism, whistleblowing and punishment for misconduct (12-16). It also seemed to us to have face validity. The questionnaire does not cover the much broader class of questionable or irresponsible research practices, but only misconduct in its strictest sense. The Kalichman questionnaire was originally designed to be used prior to teaching to generate a basis for discussion of different types of misconduct. It asks respondents to rank the level of their agreement or disagreement with each of 14 statements on a 5 point Likert scale from 'Strongly disagree' to 'Strongly agree'. Through our previous studies (12-14) we have created a data set that is sufficiently large to allow for the analysis of the reliability and factor structure of the questionnaire as a research tool.

The aim of this study is to analyse the reliability and factor structure of the Kalichman questionnaire in order to evaluate its potential future use in research on misconduct in scientific and academic communities.

Materials and methods

The data comes from three separate surveys of experiences with and attitudes towards scientific misconduct among biomedical doctoral students: 1) a 2010/11 study at the 4 Norwegian Universities having medical schools (12), 2) a 2014 study at the University of Oslo, Norway and the Karolinska Institute (KI), Sweden (13), 3) a 2015 study at the University of Oslo (14). All studies used very similar questionnaires that all included the Kalichman questionnaire. The studies are described in full in a series of papers (12-14). The questionnaires were distributed directly to doctoral students during basic compulsory courses in research methodology. The questionnaires were anonymous and participation was voluntary. We are thus unable to compare the demography of responders and non-responders. In order to emphasise the anonymous nature of the questionnaires we furthermore only collected limited demographic information to remove any suspicion in the mind of respondents that they could be deductively identified. The response rates and demographic information for the three constituent studies, and the two derived subset in this study is given in Table 1 (12-14).

The present study is a secondary analysis of anonymous data and does not require research ethics approval. The initial data collection does not require research ethics approval in Norway and Sweden since it uses fully anonymous questionnaires in a non-patient population, but the studies are registered with the Norwegian Data Protection Authority.

The reliability and properties of the questionnaire as a scale has been investigated using standard scale validation methods and confirmatory factor analysis based on Classical Test Theory (17-19). All statistical analyses have been performed using the Reliability and Factor Analysis functions in IBM SPSS version 20.

Factor analysis

For the confirmatory factor analysis the data set was split in two subsets, one containing the 2010/11, 2014 and 2015 data from the University of Oslo (the 'Oslo subset'), and one containing the 2010/11 data from other Norwegian universities and the 2014 KI data (the 'non-Oslo subset'). This method differs from the standard split-half methodology where cases are allocated randomly to

each of the two subsets. This method of splitting the dataset was chosen for two reasons: 1) it generates subsets of roughly equal size, and 2) it likely to generate the largest variation between the two subsets if, as is plausible Oslo data from one year are more similar to Oslo data from another year, than to data from other universities. The mix of students differ between the universities, as each has its specific profile. Missing data were deleted case-wise, i.e. cases were deleted if any of the 14 items were not completed. A standard exploratory factor analysis with Principal Component extraction and orthogonal Varimax rotation was performed on the Oslo subset with the number of factors extracted decided according to the Eigenvalue criterion (Eigenvalue > 1), and then a confirmatory factor analysis on the non-Oslo subset with the number of factors extracted fixed at the number extracted from the Oslo subset. Kalichman suggests that the questionnaire tap 5 domains: 1) data falsification (q1-3), 2) plagiarism (q4-6), 3) personal willingness to commit misconduct (q7-9), 4) responsibilities for whistleblowing (q10-12) and, 5) allocation of blame versus punishment (q13-14) (16). This was not reproduced in our results (see below).

An exploratory factor analysis of the total dataset also provides a 4 factor solution, so we performed an analysis with the number of factors fixed at 5 on the total dataset to investigate whether this reproduces the 5 domains suggested by Kalichman in the factor structure.

Reliability

Corrected-item total correlations and Cronbach's Alpha if included were calculated for each item. Following this one of the 14 items was removed from the scale (see below in Results). Cronbach's Alpha was also calculated for each of the sub-scales identified in the factor analyses.

Results

The total number of respondents is 467, and of those 411 (88%) completed all the 14 items in the Kalichman questionnaire. Of the 411, 216 are in the Oslo subset and 195 in the Non-Oslo subset.

The overall 14 item scale has a Cronbach's Alpha of 0.666 which increases to 0.705 if item 9 is excluded from the scale. The corrected item-scale correlation of 0.006 also indicates that item 9 does not correlate with the remainder of the items.

The detailed results of the factor analyses are shown in Table 2. The exploratory factor analysis of the Oslo subset indicates a four factor solution, with most items loading primarily on one factor only. This solution explains 66.32% of the variance. The four factors are also relatively easily interpretable as indicated by the thematic coherence of the items in each of the 4 suggested sub-scales.

The confirmatory factor analysis, with the number of extracted factors fixed at 4 shows the same pattern of factor loadings, and explains 63.1% of the variance.

The four factor solution explains 67.2 % of the variance in the total dataset

The results of the factor analysis of the total dataset with the number of factors fixed at 5 are shown in Table 3. This solution explains 71.4 % of the variance, with the last factor explaining 7.1%.

The results concerning the overall 13 item scale and the four subscales are presented in Table 4.

As a rule of thumb a Cronbach's Alpha of 0.70-0.80 is considered respectable for a scale for research use, and an Alpha over .80 as very good (17). The Alphas of the four suggested subscales are thus respectable or very good, and the Alpha for the whole 13 item scale with item 9 excluded respectable.

For an interpretation and discussion of the attitudes towards misconduct displayed by the respondents we refer the reader to our previous publications (12-14).

Discussion

Our analysis indicates that item 9 "It is more important that data reporting be completely truthful in a publication than in a grant application" correlates badly with the other items, reduces Cronbach's Alpha when included, and should be excluded from the scale. It may tap a different construct or there may be an issue with the specific phrasing of the statement. In the Kalichman questionnaire there are 5 different phrasings of the statements that the respondents are asked to evaluate, and there is poor balance between positively and negatively phrased items. This is suboptimal from a scale development point of view. The questionnaire was not originally developed as a misconduct scale, but as a teaching tool, and if good scale development practices had been followed the statement phrasings would have been more uniform and more balanced between negative and positive phrasings. The purpose of the current study is to investigate the properties of the questionnaire as it is, and not to develop a new questionnaire. The results indicate that the questionnaire is statistically reliable as a scale, despite these phrasing problems, and we have no data to support the specific rephrasing of any of the items.

A 13 item scale based on items 1-8 and 10-14 of the Kalichman "Survey 2: research misconduct" questionnaire can function reliably as a measure of the general attitude towards the acceptability of scientific misconduct. The wording of the items, however restricts the scope of use as a single scale to areas of science that produces or uses 'data'.

As mentioned above Kalichman suggests that the questionnaire tap 5 domains: 1) data falsification (q1-3), 2) plagiarism (q4-6), 3) personal willingness to commit misconduct (q7-9), 4) responsibilities for whistleblowing (q10-12) and, 5) allocation of blame versus punishment (q13-14) (16). Our findings are broadly consistent with this, although we only identify 4 factors when performing exploratory factor analysis. There is, however some support for a 5 factor solution in the magnitudes of the factor loadings for fabrication and plagiarism related items, respectively on factor 1 of the 4 factor solution. The 5 factors identified when a 5 factor solution is enforced do correspond to the 5 domains suggested by Kalichman, although the reliability of the identified Fabrication / Falsification sub-scale is not acceptable. It is a potential weakness of the study that all the data come from biomedical doctoral students at Scandinavian universities. This is, however more likely to influence the actual attitude scores than it is to influence the reliability and structure of the scale and sub-scales. The actual attitude scores furthermore indicate that there is significant variability in attitudes towards scientific misconduct even within this circumscribed population.

Conclusion

Our results indicate that Kalichman's "Survey 2: research misconduct" questionnaire is a statistically reliable tool for investigating general attitudes towards scientific misconduct. Factor analysis indicates that the overall scale can be divided into 4 reliable sub-scales each tapping a different construct related to scientific misconduct 1) general attitude to misconduct, 2) attitude to personal misconduct, 3) attitude to whistleblowing, and 4) attitude to blameworthiness/punishment for misconduct.

Research agenda

A complete investigation of the reliability and validity of the questionnaire will require further research, e.g. into criterion validity. Using the same instrument to measure attitudes towards misconduct in different studies does, however facilitate direct comparison between the findings and the growth of knowledge in the field. We hope that these results indicating that the Kalichman questionnaire is a reliable scale will facilitate the increased use of the questionnaire in scientific misconduct research. This will increase comparability between studies and may ultimately, through the generation of better data help to reduce misconduct and increase trust in researchers and in science in general.

Acknowledgements

We gratefully acknowledge the permission of our collaborators to re-use the data collected in previous studies.

We thank the reviewers for very helpful comments.

Authors' Note

The idea for this study arose in conversation between the authors. Søren Holm conducted the data analysis and wrote the first draft of the article, and both authors made substantial contributions to the revision of the article.

Declaration of Conflicting Interests

As described in the biographies the authors teach ethics to PhD students at the University of Oslo including to some of the respondents of the surveys mentioned in this study. The authors have no conflict of interests to report.

Funding

The authors received no financial support for the research, authorship or publication of this article

Author biographies

Søren Holm is a professor of bioethics at the University of Manchester and professor of medical ethics (part-time) at the Universities of Oslo and Aalborg. His main current research interests are in research ethics, the ethics and regulation of the secondary use of health data, and science ethics including misconduct and conflicts of interest.

Bjørn Hofmann is a professor of medical ethics at the Department for Health, Technology and Social Sciences, Norwegian University of Science and Technology (NTNU) Gjøvik and at the Centre for Medical Ethics at the University of Oslo. He teaches and researches in medical ethics, philosophy of science, research methodology, technology assessment, and science ethics.

Table 1 Demographic data

	Oslo 2010	Rest of Norway 2010	Oslo 2014	KI 2014	Oslo 2015		Oslo subset	Non-Oslo subset
N Returned / Distributed questionnaires	78/87	111/175	96/107	105/115	77/98		251/292	216/290
Response rate	89.7%	63.4%	88.8%	91.3%	72.1%		85.9%	74.4%
Years of doctoral study <1 / 1-2 / >2	51/17/10	67/33/11	55/28/11	61/33/10	57/13/4		163/58/25	128/66/21
Type of research Clinical/Basic/Other	31/30/16	54/24/32	75/17/13	48/41/16	45/18/10		151/65/39	102/65/48

Table 2 Reliability analysis, confirmatory factor analysis and suggested subscales

	Total dataset N=411 Cronbach's Alpha for 14 item scale = .666 Alpha for 13 item scale = 0.705				Oslo subset N=216 Original Exploratory analysis – number of factors decided by Eigenvalue >1				Non-Oslo subset N=195 Confirmatory analysis - number of factors fixed at 4					Suggested subscales with calculated Cronbach's Alpha for total dataset
Statement *In forming the scale item 7 and 8 are reverse scored*	Mean	SD	Corrected item-total correlation		Factor 1 loading	Factor 2 loading	Factor 3 loading	Factor 4 loading	Factor 1 loading	Factor 2 loading	Factor 3 loading	Factor 4 loading		
Q1. It is never appropriate to report experimental data that have been created without actually having conducted the experiment.	4.60	.914	.292		.519	.079	.191	-.065	.468	-.025	.0.38	.215		General attitude towards misconduct scale Alpha = .768
Q2. It is never appropriate to alter experimental data to make an experiment look better than it actually was.	4.76	.597	.351		.662	.153	.020	.008	.625	.112	-.197	.095		

Q 3. It is never appropriate to try a variety of different methods of analysis until one is found that yields a result that is statistically significant.	3.75	.986	.368		.549	.178	-.001	.276	.457	.0.43	-.163	.344		
Q4. It is never appropriate to take credit for the words or writing of someone else.	4.64	.678	.380		.816	-.032	-.073	.046	.724	.216	-.035	-.043		
Q5. It is never appropriate to take credit for the data generated by someone else.	4.49	.824	.438		.789	.075	-.045	.112	.850	.084	-.016	.027		
Q6. It is never appropriate to take credit for the ideas generated by someone else.	4.45	.841	.394		.846	.051	.-.087	.040	.815	.060	-.063	.016		
Q7. If you are confident of your findings, it is acceptable to selectively omit contradictory results to expedite publication.	2.03	1.207	.147		.077	-.149	.836	.037	-.096	-.048	.879	.012		Attitude to personal misconduct scale Alpha = .784
Q8. If you are confident of your findings, it is acceptable to falsify or	1.49	1.198	.224		.098	-.068	.824	0.106	-.028	.039	.850	.112		

fabricate data to expedite publication.														
Q9. It is more important that data reporting be completely truthful in a publication than in a grant application.	2.82	1.324	.006		-.157	.001	.702	-.100	-.077	-.091	.507	-.015		
Q10. If you witness someone committing research misconduct, you have an ethical obligation to act.	4.18	.789	.253		.178	.761	-.096	.091	.122	.740	-.280	.075		Whistleblowing scale Alpha = .841
Q11. If you had witnessed a co-worker or peer committing research misconduct, you would be willing to report that misconduct to a responsible official.	3.99	.747	.356		.101	.913	-.087	.091	.125	.905	.047	.070		
Q12. If you had witnessed a supervisor or principal investigator committing research misconduct, you would be willing to report that misconduct to a responsible official.	3.92	.800	.338		.076	.896	-0.35	.086	101	.902	.023	.077		

Q13. If fabricated data are discovered in a published paper, all co-authors must equally share in the blame.	3.32	1.121	.404		.078	.144	.057	.922	.124	.134	-.014	.899		Punishment scale Alpha = .877
Q14. If fabricated data are discovered in a published paper, all co-authors must get the same punishment.	2.83	1.110	.386		0.089	.089	-.023	.943	.099	.062	.151	.890		

Table 3 5 Factor solution for total dataset

	Total data set N=411 Number of factors fixed at 5					Suggested subscales with calculated Cronbach's Alpha for total dataset
Statement *In forming the scale item 7 and 8 are reverse scored*	Factor 1 loading	Factor 2 loading	Factor 3 loading	Factor 4 loading	Factor 5 loading	
Q1. It is never appropriate to report experimental data that have been created without actually having conducted the experiment.	.075	.021	.095	-.017	.845	Fabrication / Falsificationscale Alpha = .567
Q2. It is never appropriate to alter experimental data to make an experiment look better than it actually was.	.385	.096	-0.058	0.014	.613	
Q 3. It is never appropriate to try a variety of different methods of analysis until one is found that yields a result that is statistically significant.	.228	.112	-.076	.255	.614	
Q4. It is never appropriate to take credit for the words or writing of someone else.	.814	.094	-.030	.028	.153	Plagiarism scale Alpha = .847
Q5. It is never appropriate to take credit for the data generated by someone else.	.847	.106	-.020	.092	.167	
Q6. It is never appropriate to take credit for the ideas generated by	.842	.076	-.074	.054	.196	

someone else.						
Q7. If you are confident of your findings, it is acceptable to selectively omit contradictory results to expedite publication.	.029	-.103	.868	.0.35	-.053	Attitude to personal misconduct scale Alpha = .784
Q8. If you are confident of your findings, it is acceptable to falsify or fabricate data to expedite publication.	.044	-.034	.837	.127	-.010	
Q9. It is more important that data reporting be completely truthful in a publication than in a grant application.	-.176	-.035	.616	-.087	.046	
Q10. If you witness someone committing research misconduct, you have an ethical obligation to act.	.156	.755	-.171	.079	.045	Whistleblowing scale Alpha = .841
Q11. If you had witnessed a co-worker or peer committing research misconduct, you would be willing to report that misconduct to a responsible official.	.074	.916	-.022	.074	.083	
Q12. If you had witnessed a supervisor or principal investigator committing research misconduct, you would be willing to report that misconduct to a responsible official.	.051	.911	.000	.067	.073	
Q13. If fabricated data are discovered in a published paper, all co-authors must equally share in the blame.	.069	.124	.020	.915	.087	Punishment scale Alpha = .877
Q14. If fabricated data are discovered in a published paper, all co-authors must get the same punishment.	0.066	.070	.045	.931	.075	

Table 4 Total scale and subscale characteristics

	Number of items	Mean score	SD	Normalised score (mean score / number of items; Range 1 – 5)	Cronbach's Alpha
Total scale	13	53.41	5.77	4.11	0.705
General attitude scale	6	26.64	3.41	4.44	0.768
Personal misconduct scale *Items reverse scored*	2	8.50	2.18	4.25	0.784
Whistleblowing scale	3	12.12	2.01	4.04	0.841
Punishment scale	2	6.18	2.13	3.09	0.877

N = 411

References

1. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(42):17028-33.

2. Shamoo A, Resnik D. Responsible Conduct of Research, 3rd ed. New York Oxford University Press; 2014.

3. Roberts DL, St John FA. Estimating the prevalence of researcher misconduct: a study of UK academics within biological sciences. PeerJ. 2014;2:e562.

4. Bedi N. Medical research misconduct need regulatory reforms. Indian journal of community medicine : official publication of Indian Association of Preventive & Social Medicine. 2014;39(4):194-6.

5. Resnik DB, Dinse GE. Scientific retractions and corrections related to misconduct findings. Journal of medical ethics. 2013;39(1):46-50.

6. Resnik, DB, Shamoo AE. The Singapore Statement on Research Integrity. Account Res. 2011;18(2):71-75.

7. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PloS one. 2009;4(5):e5738.

8. Broome ME, Pryor E, Habermann B, Pulley L, Kincaid H. The Scientific Misconduct Questionnaire--Revised (SMQ-R): validation and psychometric testing. Accountability in research. 2005;12(4):263-80.

9. Mavrinac M, Brumini G, Bilic-Zulle L, Petrovecki M. Construction and validation of attitudes toward plagiarism questionnaire. Croatian medical journal. 2010;51(3):195-201.

10. Wester, K. L., Willse, J. T., & Davis, M. S. (2008). Responsible conduct of research measure: Initial development and pilot study. Accountability in Research, 15(2), 87-104.

11. DuBois JM, Chibnall JT, Gibbs J. Compliance Disengagement in Research: Development and Validation of a New Measure. Science and engineering ethics. 2016;22(4):965-88.

12. Hofmann B, Myhr AI, Holm S. Scientific dishonesty--a nationwide survey of doctoral students in Norway. BMC medical ethics. 2013;14:3.

13. Hofmann B, Helgesson G, Juth N, Holm S. Scientific Dishonesty: A Survey of Doctoral Students at the Major Medical Faculties in Sweden and Norway. Journal of empirical research on human research ethics : JERHRE. 2015;10(4):380-8.

14. Hofmann B, Holm S. Vitenskapelig uredelighet: Kunnskap, handlinger og holdninger hos doktorgradskandidater ved Universitetet i Oslo. Tidsskrift for den Norske laegeforening 2016; 136:1442-7.

15. Kalichman MW, Friedman PJ. A pilot study of biomedical trainees' perceptions concerning research ethics. Academic Medicine. 1992;67(11):769-75.

16. Kalichman MW. Surveys as a tool for training in scientific integrity. In: Macrina F, editor. Scientific integrity: text and cases in responsible conduct of research Washington DC: ASM Press; 2005. p. 297-320

17. DeVellis, R.F. (2017). *Scale Development – Theory and Applications (4th ed.).* Los Angeles, CA: Sage.

18. Rust, J.R.; Golombok, S. (1999). Modern Psychometrics (2nd ed.). London: Routledge.

19. Mulak, S.A. (2009). Foundations of Factor Analysis (2nd ed.). Boca Raton, FL: CRC Press.